FACULTY OF COMPUTING
SEMESTER II, SESSION 2024/2025

BACHELOR OF COMPUTER SCIENCE (BIOINFORMATICS)

SECB3203 PROGRAMMING FOR BIOINFORMATICS - SECTION 01

PROJECT
ALZHEIMER'S DISEASE PREDICTION

GROUP MEMBERS	MATRIC NO
TAN ZHAO HONG	A23CS0188
CHIN PEI WEN	A23CS0065
KOO XUAN	A23CS0300

LECTURER'S NAME : DR. SEAH CHOON SEN

SUBMISSION DATE : 03 DECEMBER 2025

Table of Contents

2.3 Software and Hardware Requirements

2.3.1 Software Tools
- Visual Studio Code
  - Integrated Development Environment (IDE) for script development and project organization
- Python
  - Programming language used for data preprocessing, analysis, and machine learning model development.
- Jupyter Notebook
  - Environment for code execution, data visualization, and reproducibility of experiments.
- Python libraries
  - Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn for data manipulation, visualization, and machine learning tasks

2.3.2 Hardware Specifications
- Computer
  - System type:64-bit operating system
  - Processor: 11th Gen Intel(R) Core(TM) i5-11300H @ 3.10GHz
  - Memory (RAM): 8GB, sufficient for handling datasets and training classical machine learning models
  - Storage: 476GB, adequate for storing datasets, model parameters, and project files

3.0 Flowchart of the Proposed Approach

```
                           ( Start )
                              |
 ┌─ Data Collection ──────────┼──────┐      ┌─ Data Preprocessing ─────────────────────┐
 │                            |      │      │                                          │
 │  ┌──────────────────────────────┐ │      │   ┌────────────────────────────────┐     │
 │  │ Importing Alzheimer's Dataset│ │─────▶│   │   Remove non-medical features  │     │
 │  │        from Kaggle           │ │      │   └────────────────────────────────┘     │
 │  └──────────────────────────────┘ │      │                  |                       │
 │                |                   │      │            ◇ Missing Values? ◇           │
 │  ┌──────────────────────────────┐ │      │        NO /              \ YES           │
 │  │  Display dataset information  │ │      │   ┌────────────────────────────────┐     │
 │  └──────────────────────────────┘ │      │   │ Fill missing values with median│     │
 │                |                   │      │   └────────────────────────────────┘     │
 │  ┌──────────────────────────────┐ │      │   ┌────────────────────────────────┐     │
 │  │   Check for missing values   │─┼─────▶│   │ Encode categorical variables    │     │
 │  └──────────────────────────────┘ │      │   │     using LabelEncoder          │     │
 └────────────────────────────────────┘     │   └────────────────────────────────┘     │
                                             │   ┌────────────────────────────────┐     │
 ┌─ Exploratory Data Analysis ───────┐       │   │ Separate features X and target y│    │
 │  ┌──────────────────────────────┐ │       │   │       Target: Diagnosis         │    │
 │  │ Analyze target distribution  │◀┼───────│   └────────────────────────────────┘     │
 │  │   Bar plot and pie chart     │ │       │   ┌────────────────────────────────┐     │
 │  └──────────────────────────────┘ │       │   │ Split data: 80% Train, 20% Test │    │
 │  ┌──────────────────────────────┐ │       │   │       Stratified split          │    │
 │  │ Generate descriptive statistics│ │      │   └────────────────────────────────┘     │
 │  │     Mean, median, std        │ │       │   ┌────────────────────────────────┐     │
 │  └──────────────────────────────┘ │       │   │ Normalize features using        │    │
 │  ┌──────────────────────────────┐ │       │   │     StandardScaler              │    │
 │  │ Create correlation heatmap   │ │       │   └────────────────────────────────┘     │
 │  │   for numerical features     │ │       └───────────────────────────────────────────┘
 │  └──────────────────────────────┘ │
 │  ┌──────────────────────────────┐ │
 │  │ Identify relationships       │ │
 │  │ between features and diagnosis│ │
 │  └──────────────────────────────┘ │
 └────────────────────────────────────┘

 ┌─ Model Development ───────────────┐       ┌─ Model Evaluation ───────────────────────┐
 │  ┌──────────────────────────────┐ │       │   ┌────────────────────────────────┐     │
 │  │    Train multiple ML models  │ │       │   │ Calculate performance metrics:  │     │
 │  │    1. Logistic Regression    │◀┼───────│   │ Accuracy, Precision, Recall,    │     │
 │  │    2. Decision Tree          │ │       │   │       F1-Score                  │     │
 │  │    3. SVM                    │ │       │   └────────────────────────────────┘     │
 │  │    4. KNN                    │ │       │   ┌────────────────────────────────┐     │
 │  └──────────────────────────────┘ │       │   │ Generate descriptive statistics │    │
 │  ┌──────────────────────────────┐ │       │   │     Mean, median, std           │    │
 │  │ Train each model on training │ │       │   └────────────────────────────────┘     │
 │  │        dataset               │ │       │   ┌────────────────────────────────┐     │
 │  └──────────────────────────────┘ │       │   │ Perform 5-fold cross-validation │    │
 │  ┌──────────────────────────────┐ │       │   └────────────────────────────────┘     │
 │  │ Make predictions on test     │─┼──────▶│   ┌────────────────────────────────┐     │
 │  │        dataset               │ │       │   │ Plot ROC-AUC curves for models  │    │
 │  └──────────────────────────────┘ │       │   │     with probability            │    │
 └────────────────────────────────────┘       │   └────────────────────────────────┘     │
                                               └───────────────────────────────────────────┘

 ┌─ Documentation ───────────────────┐       ┌─ Results Visualization ──────────────────┐
 │  ┌──────────────────────────────┐ │       │   ┌────────────────────────────────┐     │
 │  │  Record performance metrics  │◀┼───────│   │   Compare model performance    │     │
 │  └──────────────────────────────┘ │       │   └────────────────────────────────┘     │
 │  ┌──────────────────────────────┐ │       │   ┌────────────────────────────────┐     │
 │  │    Prepare final report]     │ │       │   │   Display confusion matrices   │     │
 │  └──────────────────────────────┘ │       │   └────────────────────────────────┘     │
 └────────────────────────────────────┘       │   ┌────────────────────────────────┐     │
                |                               │   │   Show ROC curves comparison   │     │
            ( End )                             │   └────────────────────────────────┘     │
                                               │   ┌────────────────────────────────┐     │
                                               │   │ Identify model with highest     │    │
                                               │   │ accuracy and Select Best Model  │    │
                                               │   └────────────────────────────────┘     │
                                               │   ┌────────────────────────────────┐     │
                                               │   │  Display final results summary │     │
                                               │   └────────────────────────────────┘     │
                                               └───────────────────────────────────────────┘
```